



COVER SHEET

This is the author-version of article published as:

Goddard, Richard and Simons, Roland and Patton, Wendy and Sullivan, Karen (2004) Psychologist hand scoring error rates on the Rothwell Miller Interest Blank: A comparison of three job allocation systems. *Australian Journal of Psychology* 56(1):pp. 25-32.

Accessed from <http://eprints.qut.edu.au>

Copyright 2004 Taylor & Francis

Psychologist hand scoring error rates on the Rothwell Miller Interest Blank: A comparison of
three job allocation systems.

Richard Goddard
School of Human Services
Griffith University

Roland Simons
Australian Centre in Strategic Management
Queensland University of Technology

Wendy Patton
School of Learning and Professional Studies
Queensland University of Technology

Karen Sullivan
School of Psychology & Counselling
Queensland University of Technology

Requests for reprints to:

Richard Goddard
School of Human Services
Griffith University Logan Campus
University Drive
Meadowbrook QLD 4131
e-mail: r.goddard@griffith.edu.au

Keywords: Rothwell Miller, Vocational Interests, Interest Inventories, Career Interest Tests,
Error Rates, Hand Scoring, Psychological Tests, Work Allocation, Time
Management, Work Environment, Reliability, Validity, Test design.

Running head: Psychologist error rates

Word Count: 4790

Psychologist hand scoring error rates on the Rothwell Miller Interest Blank: A comparison of
three job allocation systems.

Abstract

Hand scoring errors are known to occur on a range of psychological tests. The present study conducts an investigation into the existence of scoring errors by 27 professional occupational psychologists using the Rothwell-Miller Interest Blank (RMIB; Miller, Tyler & Rothwell, 1994). Building on investigations into the impact of work allocation practices on work quality in other professions, this study explored whether psychologist scoring error rates differed between three work allocation systems. Data from 1175 completed RMIB survey forms indicated error rates for the three systems ranged from five to 16.3 percent, with the self-managed work allocation system resulting in the lowest error rate. The discussion focuses on possible ways for psychologists to overcome scoring error rates with the RMIB and the potential implications these results have for allocating case work to psychologists. Suggestions for test developers and organisations designing work allocation systems are proffered.

Psychologist hand scoring error rates on the Rothwell Miller Interest Blank: A comparison of
three job allocation systems.

With the exception of a considerable body of literature focusing on professional scoring of the Wechsler scales over several decades (e.g., Franklin, Stillman, Burpeau, & Sabers, 1982; Levenson, Golden-Scaduto, Aiosa-Karpas, & Ward, 1988; Oakland, Lee, & Axelrad, 1975; Sherets, Gard, & Langner, 1979; Slate & Chick, 1989; Slate & Hunnicutt, 1988; Sullivan, 2000; Whitten, Slate, Jones, Shine, & Raggio, 1994), examiner errors which occur on other psychological tests that are hand scored has received relatively limited attention. Where research on this topic has been reported however, examiner error has been found to occur with considerable frequency and to have the potential to alter test interpretation (Allard, Butler, Shea, & Faust, 1995; Charter, Walden, & Padilla, 2000; Hunnicutt, Slate, Gamble, & Wheeler, 1990; Sherrets et al., 1979; Levenson et al., 1988; Peterson, Steger, Slate, & Jones, 1991). This study examines the susceptibility of the Rothwell Miller Interest Blank (RMIB; Miller, Tyler, & Rothwell, 1994) to hand scoring errors, since error rates on this test have not been widely investigated (for an exception see Simons, Goddard, & Patton, 2002). The RMIB is an enduring and commonly used career interest inventory that has recently been revised for Australian respondents.

The general pattern of results of previous examiner error studies suggests that a substantial proportion of scoring errors may be attributed to clerical mistakes, such as errors with transposition or calculation (Allard et al., 1995; Allard & Faust, 2000; Hunnicutt et al., 1990; Slate & Hunnicutt, 1988; Whitten et al., 1994). Whether the RMIB is susceptible to these types of error has yet to be investigated.

Importantly, a small number of studies have attempted to identify factors that contribute to scorer error. For example, the complexity of scoring procedures has been shown to be positively correlated with scoring errors (e.g., Allard & Faust, 2000; Tracey & Sedlacek, 1980; Whitten et al., 1994). A recent study by Allard and Faust (2000) found that examiner commitment to score accurately was linked to scoring accuracy. That is, examiners who had a low commitment to score accurately had significantly higher scoring error rates than examiners with high levels of commitment. In addition, variables such as test users' level of experience with particular tests has also been investigated to determine if this moderates scoring errors, however scoring errors have been reported among professional and novice scorers (Simons et

al., 2002). As yet, however, the task of identifying the full domain of variables that might explain significant human scoring error rates would seem to be far from complete.

A broader examination of the literature on human performance suggests another important determinant of work performance is the job allocation system used. Indeed, the impact of work allocation systems on performance has been the source of much debate in professions such as accountancy, engineering, law, and consulting in general (Wakefield, 2001; Weiss, 2002; Yakura, 2001), although there is little written about their impact on the work quality of psychologists. The literature on work allocation systems outside psychology describes three main methods of allocating work. These are time-based, quota-based, and self-managed. In time-based systems, the more time an employee spends on a client's job the more hours they can bill the client for. This system of job allocation, often used by accountants, was seen to have an advantage over standard billing where the total time required by any client was known to vary and where a standard fee resulted in inequities for some clients. For example, some clients paid for time that they never needed while others had their bills 'capped'. Time based billing provided added benefits for organisational management in that each bill could be tracked back to any particular staff member and the total billable hours for each staff member could be calculated. This process in turn led to the ability to determine the most and least productive employees. Limitations of time-based work allocation systems include the possibility that clients may be over billed (to account for hours) or that work is conducted at a slower rate to fill available hours. By extension, it has been argued that time-based billing does not promote efficiency; rather it promotes inefficiency. This scenario has been of particular interest to lawyers and is well documented in law journals (Syverud & Schilitz, 1999). There has been limited but interesting discussion about the ethical principals associated with this style of billing in the context of psychology (e.g., see Buono, 1999). Under this system of billing, organisations can determine the maximum number of billable hours per client.

Quota-based job allocation systems operate such that employees must reach a target allocated to them by their employer. For example, a psychologist may be required to see a specified number of clients each week to reach their quota. This system of work allocation should reward efficient work practices, but is also susceptible to short-cuts undertaken to reach the quota sooner. Under this system of work allocation, organisations can determine the number of cases that need to be seen per hour or week, for example.

Self-managed job allocation allows employees to decide how long to spend on each job, and how many jobs per week to accept. This system of job allocation allows for maximum flexibility. More complex jobs (or cases) can be allocated more time, and more or less cases can be processed depending on factors such as job complexity. Importantly, the impact of various contextual pressures on scoring accuracy has never, to the knowledge of the authors, been examined and this is the first such study. Importantly, all three job allocation systems are currently used to determine work practice in Australian psychology.

The Present Study

The present study adds to the psychological literature in two areas. First it makes a significant contribution to the existing literature on examiner scoring error by investigating scoring errors made on the RMIB using a prospective design. Until recently, examiner scoring error for this instrument has hitherto gone unreported. However a previous investigation, based on a three-year retrospective analysis of assessment data generated by six professional psychologists, has suggested that a low but significant rate of scoring errors (8%) is likely to occur over time when the RMIB is scored by professional examiners (Simons et al., 2002). Second, the present study's focus on the relationship between work allocation systems and scoring error rate investigates a contextual variable (job allocation system) which is likely to have importance for understanding professional error rates specifically, and the quality of psychological work more generally. The aim of this study therefore was threefold. First, to determine the extent to which hand-scored RMIB were susceptible to error; second to determine the impact of job allocation systems on hand scoring errors; and third to document the type of errors that occur in different work conditions.

Method

Sample

Professional psychologists who were delivering occupational psychology services directly to individual clients were invited to participate in this study. Participants were regular users of the Rothwell Miller Interest Blank (RMIB) and were employed in a variety of Australian private and public sector organisations. Psychologists were situated either in Sydney, Brisbane or Melbourne. Initially nine

psychologists who worked under a time based management system agreed to participate in this investigation. Subsequently, psychologists whose workloads were managed using a weekly quota system or who self managed their workloads were approached until nine had been recruited to each of these other two categories. In all 27 psychologists agreed to participate in the present study on an anonymous basis. Importantly participants did not know the purpose of this study was to investigate scorer error specifically. Participating psychologists had a mean age of 31.8 years ($SD = 7.52$) and an average of 6.74 years experience ($SD = 1.60$) as professional psychologists.

Six psychologists using the time based management approach to managing their workload indicated that they were required to account for their time using 5-minute accounting intervals. The remaining three psychologists in this category used 15-minute accounting periods. Psychologists in this category reported conducting between 18 and 20 one-hour consultations with individual clients each week.

Nine psychologists managed their workloads using a weekly quota system. Under this system psychologists were required to achieve a minimum of 18 consultations each week, however within this limitation they could manage their workflow as the situation required.

Nine psychologists indicated that they were able to control their own workload allocations and set the number and length of consultations with individual clients themselves. All nine psychologists in this category agreed that they were working under a self-managed work allocation system, and reported undertaking 15 to 20 consultations each week.

All 27 psychologists indicated that they had little control over who the actual clients that they saw would be, and all psychologists, except one, were salaried employees of large organisations that provided professional occupational services to the Australian public. One psychologist worked for a large organisation on a subcontracting basis. No significant differences in average age or length of professional experience were found between the three groups of psychologists investigated in the present study.

Instrument

Originally developed in the 1950s, the RMIB was designed as a practical aid for career counsellors and others who needed a focus for a career guidance interview. A comparative measure that examines respondents' interest across twelve fields of work, the instrument was designed to be "simple and

quick” for respondents (Miller, Tyler, & Rothwell, 1994, p. ix). The technical manual indicates that the instrument is suitable for individual or group administrations (Miller et al., 1994, p. 8), and can be hand scored by both professional users and adult respondents who are suitably supervised (Miller et al., 1994, p. 11). Respondents are asked to rank, in order of preference, nine sets each of twelve jobs representing the above fields of work. The instrument takes approximately 15 minutes to complete and scoring is done on the form itself, following the procedures laid out in the manual. No mention of scoring error is made in the manual, although an arithmetical method to check computational accuracy is described (Miller et al., 1994, p. 11).

Procedure

Psychologists were directed to continue to use the RMIB with their clients as they deemed appropriate. In all cases this involved hand scoring the instrument themselves. Data were collected monthly over a three-month period during 2000. On three occasions, at the end of each month, participating psychologists sent all RMIB questionnaires that had been completed during that month directly to the researchers. In all cases the RMIB had been administered to each client during an individual consultation, that is, not in a group format. Overall, 27 psychologists supplied a total of 1175 completed RMIB survey forms to the researchers.

Analysis

Client responses from each RMIB were entered into a database twice and compared for accuracy using a computer algorithm. After obtaining concordance between the two response sets these data were then computer scored using a scoring algorithm for the RMIB that had been calculated and checked for accuracy previously. Note that use of computer based scoring algorithms is a procedure recommended by Allard et al. (1995) to enhance scoring accuracy. Records of psychologist scores derived from scoring sheets were then checked for accuracy against computer scores generated electronically. Where a discrepancy between the computer and hand scored data was detected, the original RMIB and corresponding psychologist scoring details were examined to ensure errors were genuine. In this way error

rates were identified, firstly by computer re-scoring, followed by confirmation of the error by human inspection of the original questionnaire and psychologists' scoring of this.

Confirmed errors were then classified by type into one of three primary categories, following precedents set by other researchers (e.g., Sullivan, 2000). The three primary error categories were: calculation, transcription, and total errors. Errors occurring as the result of a mistake when performing a calculation were coded as calculation errors. Errors resulting from a mistake in transcribing a score or sub-total were classified as transcription errors. Total errors were the number of calculation and transcription errors. These errors represent lower-order or simple errors.

A fourth category of error was also calculated for the purpose of this study. These were profile errors; defined as errors that resulted in misclassification of the client's area of interest or more specifically, as a profile of ranking scores for the RMIB that was different (in any way) from what would have been achieved had no error been made. Note that profile errors were not independent of other error types, since an error at the point of profiling may have been an artefact of previous transcription or calculation errors. These errors were included in this study as an indicator of the potential impact of errors on psychologists' recommendations however, and represent a higher-order error.

Finally, for some analyses an "error rate" was calculated. As not all calculation errors could be regarded as independent (i.e., basic arithmetical errors could be compounded when total scores were computed), we adopted a conservative criterion when identifying error for these calculations and classified protocols as either error free or containing one or more error. Error rate was then defined as the number of protocols with one or more error divided by the number of protocols. Note that a conservative approach to defining errors is not uncommon practice (e.g., see Whitten et al., 1994), and that this particular definition of error rate has been used elsewhere (Charter et al., 2000; Simons et al., 2002), therefore selected analyses were calculated using this definition of error.

Results

The results of this study are presented in two parts. The first set of analyses was undertaken to characterise the nature and extent of RMIB scoring errors respectively. The second set of analyses was

undertaken to determine the extent to which job allocation systems impact scoring error. All tests of statistical significance were calculated using an alpha level of 0.05, unless otherwise stated.

Characterising the nature and extent of RMIB scoring errors

To investigate the extent of RMIB hand scoring errors, descriptive statistics were calculated across forms, then persons. These analyses were undertaken to determine the distribution of errors across forms (i.e., the percentage that contained errors) and persons (i.e., the percentage of psychologists that made errors). Analyses conducted to examine the distribution of errors across scorers revealed that most psychologists made errors (25/27). This suggests that errors were made by most participants and were not restricted to a small number of psychologists. Analyses conducted across forms suggest that most forms (i.e., 88.35 percent of forms) were error-free. Conversely, approximately one in 10 RMIB forms contained at least one hand scoring error. Of those forms that contained error, the average number of errors (mode) per form was three. Further, on forms with one or more error the frequency of transcription versus calculation errors was similar (i.e., 117 versus 112 errors respectively). Overall, the results of these analyses suggests where errors were made they were distributed across forms (i.e., not confined to a small number of very error-prone forms), distributed across scorers (i.e., not confined to a small number of psychologists), and that both calculation and transcription errors occurred with similar frequency.

Impact of job allocation system on scoring error.

To investigate whether job allocation systems impact the RMIB overall hand scoring error rate, a one-way ANOVA was calculated. The dependent variable for this analysis was the total number of errors on RMIB protocols. The independent variable (group) had three levels; each corresponding to a different job allocation system. The results of this analysis revealed a significant difference in the number of RMIB hand scoring errors between groups, $F(2, 1172) = 6.146, p = .002$. Post-hoc analyses revealed significantly fewer total hand-scoring errors in the self-managed system ($M = .33; SD = 1.45$) than in the time- or quota managed job allocation systems (time-based $M = .74; SD = 1.89$; quota-based $M = .64; SD = 1.79$). The error rates for time- and quota-based systems were not significantly different from each other.

To determine if similar results would obtain for specific error types, two additional ANOVAs were run. The dependent variable in the first analysis was the number of transcription errors per RMIB form. The dependent variable in the second analysis was the number of calculation errors per RMIB form.

The pattern of results for calculation errors was similar to that obtained for the total number of errors. Specifically, the results of the overall ANOVA and subsequent post-hoc comparisons showed that significantly fewer calculation errors were made by psychologists in the self-managed condition ($M = .19$; $SD = .92$), than in either of the other conditions, which did not differ from each other, $F(2, 1172) = 4.552$, $p = .011$ (time-based group $M = .41$; $SD = 1.22$; quota-based group $M = .38$, $SD = 1.19$). The pattern of results for transcription errors was different from that obtained for the other two error comparisons. That is, the results of the overall ANOVA and post-hoc comparisons showed that whilst significantly fewer transcription errors were made in the self-managed job allocation condition ($M = .14$; $SD = .633$) than the time-based condition ($M = .33$; $SD = .90$), no other group differences were significant though trends in the expected direction were apparent, $F(2, 1172) = 5.921$, $p = .003$ (Quota-based $M = .27$; $SD = .80$).

Figure one illustrates the main findings from analyses conducted to determine the impact of job allocation system on error rates. This figure shows the absolute number of errors per form for two error types (transcription and calculation errors) plus the total number of errors. Figure one shows that significantly fewer calculation and total errors were made in the self-managed system than either the time- or quota-based job allocation conditions, which were not significantly different from each other. Significantly fewer transcription errors were made in the self-managed condition than the time-based job allocation condition, and this was the only significant group difference for this error type.

Insert Figure 1 about here

Finally, to determine the impact of job allocation system on profiling errors calculated to demonstrate the potential impact of lower order errors, a one-way ANOVA was conducted. For this analysis a categorical dependent variable was created according to whether incorrect RMIB protocols ($n = 137$) included a profiling error (88 protocols) or not (49 protocols). The independent variable for this analysis was job allocation system with three levels. The results of this analysis revealed significant group differences, $F(2, 134) = 21.181$, $p = .000$. Post-hoc comparisons revealed that significantly fewer RMIB protocols contained profiling errors when workflow was self-managed ($n = 22$), compared to other workflow management systems (quota-based system $n = 50$; time-based system, $n = 65$).

Discussion

The primary aim of this study was to determine the extent to which hand-scored RMIB protocols were susceptible to error. Previous studies have demonstrated that other psychological tests are susceptible to hand-scoring error, however little is known about the extent to which such errors occur on the RMIB. A secondary aim of this study was to investigate the impact of hand-scoring errors on RMIB protocols under three job allocation conditions. The three job allocation conditions that were compared were self-managed, time-based or quota-based systems. Importantly, this study is the first of its type to examine the impact of work context on hand-scoring error rates for any psychological test. Finally, the third aim of this study was to document the type of errors that were made and determine if different types of errors occur in different job allocation conditions.

Consistent with previous research that has demonstrated that selected psychological tests may be susceptible to clerical scoring errors (e.g., Allard et al., 1995; Allard & Faust, 2000; Hunnicutt et al., 1990; Whitten et al., 1994), the results of this study suggest the RMIB is also susceptible to such error. Specifically, our results show that the mean RMIB hand-scoring error rate (defined as the number of protocols with one or more error divided by the total number of protocols) ranged from five percent to 16.3 percent across all psychologists in this sample. This finding is consistent with a recent study that showed a mean error rate of more than five percent when professional psychologists working in the employment services sector scored the RMIB (Simons et al., 2002). Taken together, both studies on the susceptibility of the RMIB to hand scoring error suggest that *at least* one in twenty RMIB protocols scored by psychologists working in the Australian employment services sector may contain errors.

This raises questions about whether this level of error is acceptable and if not, what might be done to reduce it. Defining an acceptable level of error firstly requires investigation of the nature and extent of errors made on the RMIB. Such information may also inform error prevention strategies. In this study, three primary error types were identified: calculation, transcription, and total errors per protocol, as well as a measure of the potential impact of lower order error on professional decision making (profiling errors). The number of calculation and transcription errors that occurred was similar for each group. Where significant group differences were identified these tended to reflect fewer calculation or transcription errors

in the self-managed job allocation condition than either of the two other job allocation conditions. One way of determining if this level of error is acceptable is to determine how often errors impacted on psychologists' decision-making and recommendations. Whilst not assessed directly, the results from this study suggest that of those protocols that contained lower order errors ($n = 137$) approximately 64 percent of these contained profiling errors which may have resulted in incorrect conclusions being drawn from this test. Further, significant groups differences were found for this error type, with fewer profiling errors identified for psychologists who managed their own workflow. Overall, these findings suggest that transcription and calculation errors contribute to a substantial number of profiling errors, and that error reduction strategies need to be considered.

It is noteworthy that the RMIB scoring system appears to involve a relatively uncomplicated method of scoring, one that has been described by the test authors as "quick and simple" (Miller et al., 1994, p. ix). The complexity of calculations required could be considered low (i.e., require simple addition of numbers between one and 12), however calculations are repetitious and depend on considerable transcription of scores from one section to another. Quick, simple and repetitious calculations in combination with significant transcription requirements may be prone to lapses of concentration. Whilst respondents were not asked to indicate whether they had employed the checking procedure outlined in the current user's manual for the RMIB (Miller et al., 1994, p. 11), it was observed that many errors could have been detected if the checking procedure had been carried out as recommended. The evidence here strongly supports a call for scorers to always employ the scoring checking strategy as a part of the overall scoring procedure for this instrument. The inclusion of an additional row for checking at the bottom of the scoring sheet may facilitate checking and future research on the effects of modified form design may identify ways of reducing error and prompting users that checking transcription of data and subsequent calculations was an integral part of the scoring process.

Modifying the job allocation system that psychologists work under may be another way of reducing RMIB hand scoring errors. That is, the results of this study suggest that hand scoring errors are likely to be reduced by working in a self-managed environment, rather than under time- or quota-based job allocation systems. This finding is consistent with other studies demonstrating that the way workload is

managed can significantly impact on the quality of work produced (e.g., Semler, 1993; Spector, Dwyer, & Jex, 1988).

Importantly, the findings of this study extend the recent work of Allard and Faust (1998), which attempted to identify variables which may account for significant human scoring error rates on psychological tests. Specifically, this study is the first published work to explore the impact of psychologists' work environments in relation to scoring error. In Australia, psychologists in the employment services sector generally operate within a framework of extremely demanding workloads (Goddard, Patton, & Creed, 2000, 2001), and work pressure has consistently been linked to performance in the literature on job stress and burnout (Beehr & Newman, 1978; Spector et al., 1988). The data from the present study may reflect this link by identifying significant workload differences inherent in or arising from the differing systems of allocating workloads. Perhaps a time based system which treats every five or fifteen minute interval as an equivalent unit of work results in the experience or perception of greater work pressure for professional psychologists undertaking complex human service work, than does a self-managed work allocation system which might allow greater flexibility and appears to result in more accurate scoring of protocols. Clearly the results of this investigation pose some interesting questions for organisational researchers, as well as practitioners, regarding the effect that different work allocation and time management strategies can have on the quality of service outcomes for psychologists. Future studies could look at examining the effect of such work allocation systems on other measures of work quality by psychologists and other human service workers.

The general finding that the quality of work produced by psychologists operating under different work allocation systems varied significantly, in relation to scoring errors, warrants further investigation. The impact of time-based billing on employee work quality was suggested to be a function of pressure to increase work rate, however this link needs further examination. It is interesting to note that self-reports from the 27 psychologists indicated that despite different workload management arrangements, the number of weekly RMIB client consultations reported for each workload system was comparable. However, as an objective measure of overall client throughput was unavailable, the explanation that real differences in work pressure can explain the observed error rate differences could not be investigated. Furthermore, as equity theory is based on perceptions of investments and returns (Adams, 1965), the explanation that the

perception of higher work pressures can lead to significantly higher error rates cannot not be discounted. Finally, it should be noted that there are other potential relationships, leading to different explanations of the scoring errors, which will need to be investigated before the results of the present study can be fully integrated with theory. For example, the most notable impact of time-based billing on psychologists may not be to increase the through-put of their clients but rather to increase the number of tests they run with any particular client. Such a process of value adding could be a result of market competition for business. In effect, organisations could be arguing that you “get more for your dollar”. In summary then, whilst the argument that psychologists working under time-based billing may process significantly more cases cannot be supported by the data available from the present study, a clear difference between the error rates of the various systems has been identified.

Two important limitations of the present study are that it has only focused on the professional scoring of one instrument, the RMIB, and that the number of psychologists investigated is small. However, whilst the results of the present study must remain tentative and need to be replicated, they highlight potentially serious issues for test designers and test users. For test designers, the study adds to an earlier call for data on error rates to be routinely presented and discussed in test manuals and user guides (Simons et al., 2002), in much the same way as Holland, Powell and Fristzsche (1994, p. 15-16) have discussed hand scoring error rates that can be expected from clients self-scoring the Self-Directed Search. In the light of growing evidence of significant rates of examiner scoring error, it is may no longer be acceptable for test designers to avoid a detailed analysis of the likely incidence of hand scoring errors in applied settings. From the point of view of a professional practitioner, the purchase of the most appropriate test for a given situation may be facilitated by the inclusion of information regarding the likelihood of scoring errors in professional user manuals (or at least points of vulnerability). Finally, from the point of view of organisations managing the workflow of professional psychologists, the present study presents evidence of a clear link between work allocation system and the quality of professional work in one of the most important areas that the profession operates. Despite the limitations and difficulties of conducting prospective research into such a sensitive area for professionals clearly both the profession and the industry have a responsibility to clarify this link by conducting further research into employment practices currently operating within the profession.

References

- Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (pp. 267-299). New York: Academic Press.
- Allard, G., Butler, J., Shea, M. T., & Faust, D. (1995). Errors in hand scoring objective personality tests: The case of the Personality Diagnostic Questionnaire-Revised (PDQ-R). Professional Psychology: Research and practice, 26(3), 304-308.
- Allard, G., & Faust, D. (2000). Errors in scoring objective personality tests. Assessment, 7(2), 119-131.
- Beehr, T. A., & Newman, J. E. (1978). Job stress, employee health, and organisational effectiveness: A facet analysis, model, and literature review. Personnel Psychology, 31, 665-699.
- Buono, A. F. (1999). The ethical practice of Psychology in organizations. Personnel Psychology, 52 (2), 503-506.
- Charter, R. A., Walden, D. K., & Padilla, S. P. (2000). Too many simple scoring errors: The Rey Figure as an example. Journal of Clinical Psychology, 56(4), 571-574.
- Franklin, M. R., Stillman, P. L., Burpeau, M. Y., & Sabers, D. L. (1982). Examiner error in intelligence testing: Are you a source? Psychology in the Schools, 19, 563-569.
- Goddard, R.C., Patton, W., & Creed, P. (2000). Case manager burnout in the Australian Job Network. Journal of Applied Health Behaviour, 2(2), 1-6.
- Goddard, R.C., Patton, W., & Creed, P. (2001). Psychological distress in Australian case managers working with the unemployed. Journal of Employment Counseling, 38, 50-61.
- Holland, J. L., Powell, A. B., & Fritzsche, B. E. (1994). The Self-Directed Search (SDS) Professional User's Guide 1994 Edition. Odessa, FL: Psychological Assessment Resources.
- Hunnicut, L. C., Slate, J. R., Gamble, C., & Wheeler, M. S. (1990). Examiner errors on the Kaufman Assessment Battery for Children: A preliminary investigation. Journal of School Psychology, 28, 271-278.
- Levenson, R. L., Golden-Scaduto, C. J., Aiosa-Karpas, C. J., & Ward, A. W. (1988). Effects of examiners' education and sex on presence and type of clerical errors made on WISC-R protocols. Psychological Reports, 62, 659-664.

- Miller, K. M., Tyler, B., & Rothwell, J. W. (1994). Rothwell Miller Interest Blank Manual (Revised Australian Edition). London: Miller & Tyler.
- Oakland, T., Lee, S. W., & Axelrad, K. M. (1975). Examiner differences on actual WISC protocols. Journal of School Psychology, 13, 227-233.
- Peterson, D., Steger, H. S., Slate, J. R., & Jones, C. H. (1991). Examiner errors on the WRAT-R. Psychology in the Schools, 28, 205-208.
- Semler, R. (1993). Maverick! The success story behind the world's most unusual workplace. London: Century.
- Sherrets, S., Gard, G., & Langner, H. (1979). Frequency of clerical errors on WISC protocols. Psychology in the Schools, 16(4), 495-496.
- Simons, R., Goddard, R. C., & Patton, W. (2002). Hand scoring error rates in psychological testing. Assessment, 9(3), 292-300.
- Slate, J. R., & Chick, D. (1989). WISC-R examiner errors: Cause for concern. Psychology in the Schools, 26, 78-84.
- Slate, J. R., & Hunnicutt, L. C. (1988). Examiner errors on the Wechsler Scales. Journal of Psychoeducational Assessment, 6, 280-288.
- Spector, P. E., Dwyer, D. J., Jex, S. M. (1988). Relation of job stressors to affective, health, and performance outcomes: A comparison of multiple data sources. Journal of Applied Psychology, 73, 11-19.
- Sullivan, K. (2000). Examiners' error on the Wechsler Memory Scale-Revised. Psychological Reports, 87, 234-240.
- Syverud, K. D., & Schiltz, P. J. (1999). On being a happy, healthy, and ethical member of an unhappy, unhealthy, and unethical profession. Vanderbilt Law Review, 52(4), 869-951.
- Tracey, T. J., & Sedlacek, W. E. (1980). Comparison of error rates on the original Self-Directed Search and the 1977 revision. Journal of Counseling Psychology, 27, 299-301.
- Wakefield, R. L. (2001). Service quality. The CPA Journal, 71(8), 58-60.
- Weiss, A. (2002). Consultants work far too hard and that's nothing to be proud of. Consulting to Management, 13(3), 14.

- Whitten, J., Slate, J. R., Jones, C. H., Shine, A. E., & Raggio, D. (1994). Examiner errors in administering and scoring the WPPSI-R. Journal of Psychoeducational Assessment, 12, 49-54.
- Yakura, E. K. (2001). Billables: The valorization of time in consulting. The American Behavioral Scientist, 44(7), 1076-1095.

Figure 1. Mean number of errors by error type and job allocation condition. Standard error bars for each group are shown ($n = 1175$).

